Assignment 1

SEM 2: Structural Equation Modeling

Please hand in a `.pdf` file containing your report and a `.R` containing your codes or screenshots of every Jasp analysis. The deadline of this assignment is Tuesday May 15 15:00.

Assignment

Part 1: Expected values

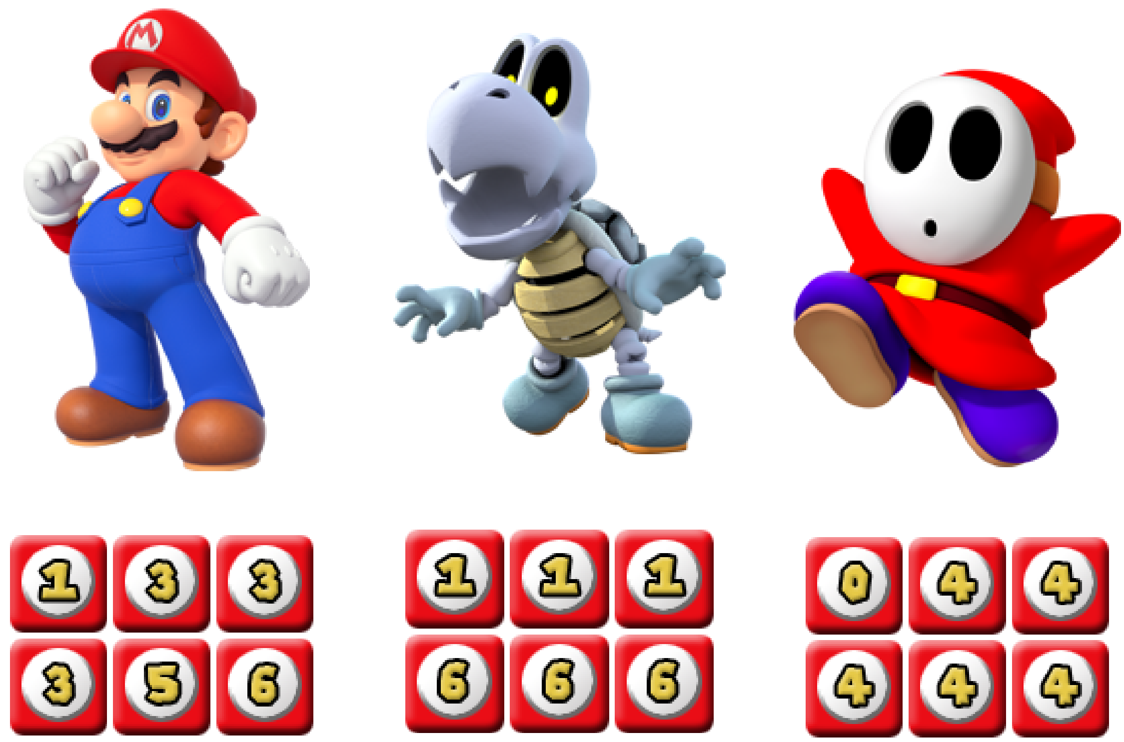
I often play the game *Super Mario Party* on the Nintendo Switch:

In this game, you play a virtual board game with three friends. Each player plays a different character. The game unfolds over a set of turns, which start by trowing a die:

After you throw a die, you move the amount of spaces equal to the number you threw. Typically, it is good to throw high numbers and move many spaces, as it will get you faster to your goal than your opponents. However, sometimes it is also good to throw a specific number (e.g., throw a 4 to land on a specific spot). As with many casual games, it helps to get an edge in the game by using math! Note, for all questions in this section, use the formulas for the expected values and variances as explained in the video lecture. Do not just give the answer, show how you obtained it in your PDF report.

Question 1 (1 point)
The standard die contains 6 values: $1,2,3,4,5,6$, which are all equally likely. What is the *expected* number of steps you can move when throwing the standard die (0.5 point) and what is its *standard deviation* (0.5 point)? Note: the standard deviation is the square root of the variance.

In the game, you choose a character to play. This has some consequences. Most importantly, each character also has a *unique* die specific to that character. At the start of your turn, you can choose to throw either the standard die or the character specific die. The dice of three characters are the following:



(a) Mario (b) Dry Bones (c) Shy Guy

Question 2 (2 points)
For each of these characters, calculate the expected number of steps you can move when throwing their specific die (1 point) and the *standard deviation* of this number of steps steps (1 point).

A concept related to variance is the *Shannon entropy*, which is defined as the expected value of the negative log likelihood:

$$\mathrm{H}(X) = \mathscr{E}(-\log_2(\mathrm{P}(X = x))).$$

Here $\log_2$ indicates the logarithm with base 2 (in R: `log(..., base = 2)`), X indicates a random outcome (e.g. a dice result 1, 2, 3, 4, 5, or 6) and $\mathrm{P}(X = x)$ indicates the probability that X takes on some value (e.g., in a standard dice $1/6$ for every possible outcome). Like the variance, the above expression reads as an expected value of a function of a random variable, and we can readily calculate this when we know the full probability distribution. For example, for a standard die we obtain:

```
1/6 * -log(1/6, 2) + # Outcome = 1
  1/6 * -log(1/6, 2) + # Outcome = 2
  1/6 * -log(1/6, 2) + # Outcome = 3
  1/6 * -log(1/6, 2) + # Outcome = 4
  1/6 * -log(1/6, 2) + # Outcome = 5
  1/6 * -log(1/6, 2) # Outcome = 6

## [1] 2.584963
```

This value gives the average number of bits (0 or 1 in computers) needed to encode the outcome[1]. A lower entropy means that less outcomes are likely, regardless of the value of that outcome. For example, if we flip a coin and get €1,000 if we get heads but pay €1,000 if we have tails, the expected value of the money we win is €0, its variance and standard deviation are high (each flip we win or lose a lot of money), but the Shannon entropy is low (there are only two outcomes: heads or tails).

Question 3 (1 point)
Calculate the *Shannon entropy* for the dice of every character.

We now have three measures per die: (1) the expected value encoding the average amount of steps we move when throwing a die, (2) the variance/standard deviation encoding the spread (uncertainty) around that expected value, and (3) the Shannon entropy encoding the uncertainty in number of outcomes, but not the amount of steps. We also have *four* dice: the standard die you investigated in question 1, and the character specific dice you investigated in questions 2 and 3.

Question 4 (1 point)
Which die do you think is the most appropriate in the following settings? Briefly argue your answer (at most one sentence)

- I want to throw a die that gives me the most predictable outcome
- I want to move the most amount of steps on average
- I want to move a lot of steps on average, but also want to be consistent in this (at least move a few steps every turn)

[1] One interpretation is that the average number of yes/no questions we need to answer lies between $\mathrm{H}(X)$ and $\mathrm{H}(X)+1$. For example, if we want to now the result of a fair die, we can ask the questions (1) is the result above 3? and (2) is the result even? Now, if the result is 2 or 5 we already know the answer, but if it is not we need to ask one more question to distinguish between 1 and 3 or 4 and 6. As this is twice as likely, we would have to ask this question $2/3$ times. Thus, on average, we need to ask $2.666\ldots$ questions, which is between $\mathrm{H}(X)$ and $\mathrm{H}(X)+1$. See also `https://math.stackexchange.com/questions/2916887/shannon-entropy-of-a-fair-dice`

In the game, you can also collect and use some *items*. For example, you can obtain the *golden dash mushroom*:

The effect of the golden dash mushroom is that a fixed score of 5 is added to your die throw, making you move 5 steps further.

Question 5 (1 point)
Choose your favorite die, and calculate the expected number of steps you would move as well as the standard deviation after using a golden dash mushroom. Use the rules of expectation and covariance algebra as discussed in the video lecture.

Another mechanic in the game is that you may at some point gain an *ally*:

The effect of an ally is that every time you throw a die, your ally also *independently* throws a small die that has a 50% chance of throwing a 1 and a 50% chance of throwing a 2. Your throw is summed with your allies throw to give you the total number of steps you move. So, for example, if you throw a 4, and your ally throws a 1, you move $4+1=5$ steps.

Question 6 (1 point)
Choose your favorite die, and calculate the expected number of steps you would move as well as the standard deviation when you have an ally. Use the rules of expectation and covariance algebra as discussed in the video lecture.

Question 7 (1 point)
Conceptually, describe if you think the Shannon entropy changes if you (a) use a golden dash mushroom (0.5 point) or (b) use an ally (0.5 point).

Part 2: Covariance algebra

Given the following structural equations:

$$
\begin{aligned}
y_{i1} &= \beta_{21} y_{i2} + \beta_{31} y_{i3} + \varepsilon_{i1} \\
y_{i2} &= \beta_{42} x_i + \varepsilon_{i2} \\
y_{i3} &= \beta_{43} x_i + \varepsilon_{i3}
\end{aligned}
$$

Question 8 (2 points)
Draw the corresponding path diagram. Make sure the path diagram contains every parameter (1 point), is clearly labeled (0.5 point), and is clear (0.5 point).

Question 9 (2 points) How many degrees of freedom does this SEM model have?

Question 10 (1 point) Derive $\text{Var}(y_1)$ using (univariate) covariance algebra.

Question 11 (1 point) Derive $\text{Cov}(y_1, x)$ using (univariate) covariance algebra.

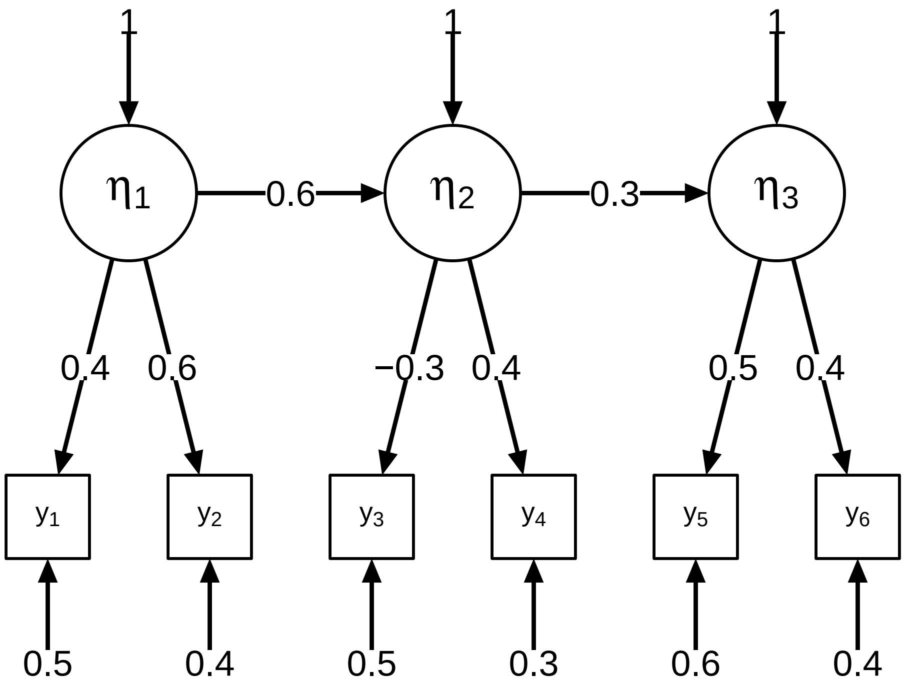
Question 12 (1 point) Derive $\text{Cov}(y_2, y_3)$ using (univariate) covariance algebra.

Question 13 (1 point) Confirm your answers to questions 10 to 12 using Wright's path tracing rules.

Part 3. Modeling SEMs

Consider the following model:

```
## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang
## Registered S3 methods overwritten by 'huge':
##   method    from
##   plot.sim  BDgraph
##   print.sim BDgraph
```



Question 14 (1 point) Give the matrices $\boldsymbol{\Lambda}$, $\boldsymbol{B}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ (containing values, not symbols). ▪

Question 15 (1 point) Compute $\boldsymbol{\Sigma}$ using the all-y formula from the slides. Note that $\boldsymbol{I}$ is an identity matrix, a matrix with ones on the diagonal and zeroes elsewhere, of the same dimensions as $\boldsymbol{B}$. ▪

Question 16 (1 point) Use Wright's path tracing rules to determine $\mathrm{cov}(y_1, y_6)$, and compare your answer to Question 15. ▪

It is known that $(\boldsymbol{I}-\boldsymbol{B})^{-1}$ can be written as:

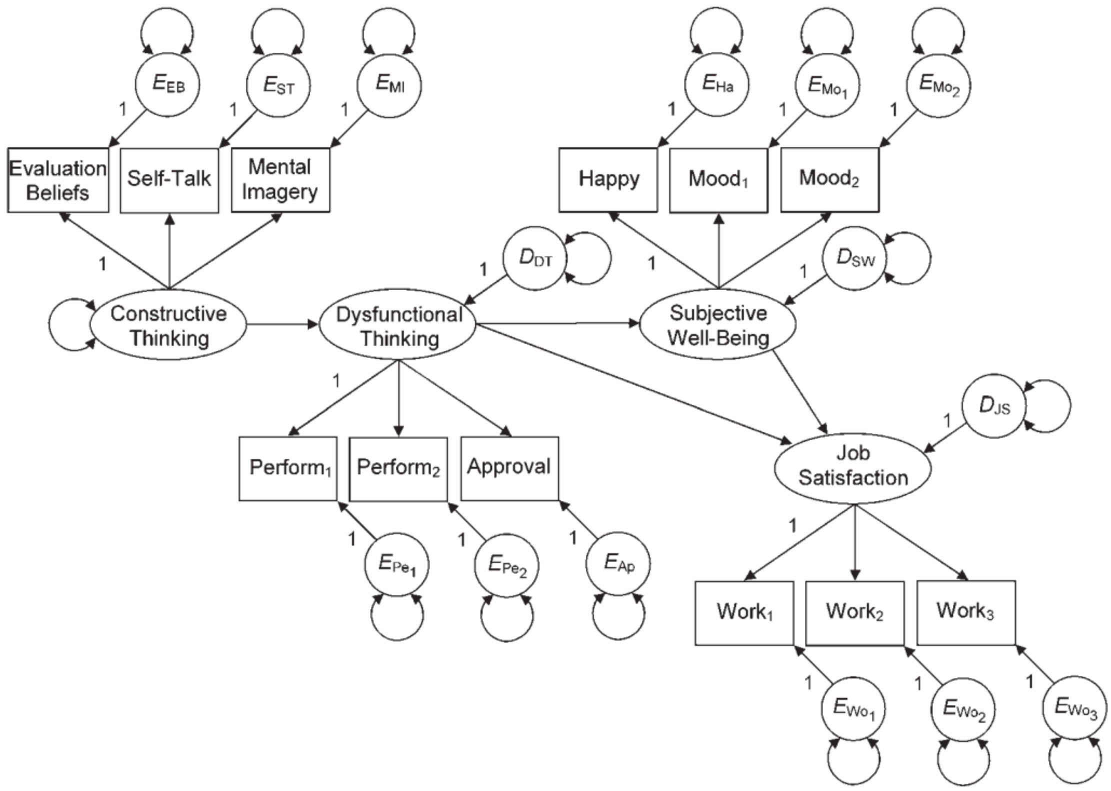
$$\begin{aligned}(\boldsymbol{I}-\boldsymbol{B})^{-1} &= \boldsymbol{I}+\boldsymbol{B}+\boldsymbol{B}^2+\boldsymbol{B}^3+\boldsymbol{B}^4+\dots \\ &= \boldsymbol{I}+\boldsymbol{B}+\boldsymbol{B}\boldsymbol{B}+\boldsymbol{B}\boldsymbol{B}\boldsymbol{B}+\boldsymbol{B}\boldsymbol{B}\boldsymbol{B}\boldsymbol{B}+\dots\end{aligned} \tag{1}$$

if $\boldsymbol{B}$ is *nilpotent*, meaning that for some finite k, $\boldsymbol{B}^k = \boldsymbol{O}$ (a matrix containing only zeroes) and as a result for every $l \in \mathbb{Z}^+$ (l is a positive integer, such as 1, 2 or 3), $\boldsymbol{B}^{k+l} = \boldsymbol{O}$.

Question 17 (1 point) Show that $\boldsymbol{B}$ is nilpotent for $k \geq 3$ and compute $(\boldsymbol{I} - \boldsymbol{B})^{-1}$ using Equation (1). Also compute $(\boldsymbol{I} - \boldsymbol{B})^{-1}$ using R and compare your results. ■

Part 4. Fitting SEMs

Kline (2015) reports a path diagram of a SEM analysis that was originally performed by Houghton and Jinkerson (2007):



Question 18 (1 point) Verify that the degrees of freedom in this model are 50 (note: the original authors report DF = 51, which seems to result from some mistake made in the analysis).

The original article contains a correlation matrix, standard deviations, and the sample size ($N = 263$). With this information, we can also construct the variance–covariance matrix, which I prepared for you in the file `houghton.csv` on Canvas, which can be loaded in R as follows:

```
covMat <- as.matrix(read.csv("houghton.csv"))
rownames(covMat) <- colnames(covMat)
```

Question 19 (2 points) Fit the SEM model to the data (1 point), and judge the fit of your model (1 point). Note, the original article will report something different, likely due to the same mistake that led to the wrong degrees of freedom.

Question 20 (1 point) We may be interested to test if subjective well-being fully mediates the effect of dysfunctional thinking on job satisfaction. To this end, we can fix the direct path from dysfunctional thinking to job satisfaction to zero, and compare the fit of the restricted model to the general model you tested above. Can you conclude that subjective well-being fully mediates the effect between dysfunctional thinking and job satisfaction?

References

Houghton, J. D., & Jinkerson, D. L. (2007). Constructive thought strategies and job satisfaction: A preliminary examination. *Journal of Business and Psychology*, *22*(1), 45–53.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.